MODULATION IN A MOBILE TELECOMMUNICATIONS SYSTEM

Field of the Invention

The present invention relates to mobile telecommunications; in particular, to a method of communication of data in a mobile telecommunications network, to a mobile telecommunications network, to a transmitter and to a receiver.

The invention was made in the course of work relating to multiple-input multiple-output (MIMO) telecommunications systems, but the invention can relate to other telecommunications systems.

Description of the Related Art

Multiple-input multiple-output (MIMO) techniques are well known, and the reader is referred to, for example, G. J Foschini and M. J. Gans "On limits of wireless communications in a fading environment when using multiple antennas", Wireless Personal Communications, vol. 6, pp. 311-335, 1998, as background. MIMO radio links have been suggested for use in code division multiple access (CDMA) networks, such as Universal Mobile Telecommunications System (UMTS) telecommunications networks in particular with high-speed downlink packet access (HSDPA) schemes. The underlying idea of HSPDA is to increase the achievable data rates for a particular user through a combination of spreading code re-use across transmit antennas and higher-order modulation schemes. However, the code re-use inevitably results in high levels of interference at the mobile receiver, even under non-dispersive channel conditions.

In order to tackle such high interference levels, MIMO receivers based on the *a posteriori* probability (APP) detector have been proposed. In order to deal with dispersive channels (and hence to avoid sequence estimation) it is necessary to precede such an APP detector with a space-time channel equalizer, followed by a de-spreading operation which allows the APP to perform joint detection of bits transmitted from multiple antennas but corresponding to a single spreading code only, thereby resulting in a significant reduction in computational complexity.

More recently, a multi-stage partial parallel interference canceller (MS-PPIC) has been proposed as an alternative to the APP detector within the above-described receiver structure. Such interference cancellation (SIC) schemes have been considered for many years in the context of multi-user detection for the CDMA uplink.

When using high-order modulations, known MIMO receivers experience problems. For example, the MS-PPIC based detector is manageable in complexity, but provides poor performance for higher order modulations. On the other hand, the APP detector becomes too complex to implement due to its exponential growth in computational complexity.

Specifically, in the known MIMO receiver based on an APP detector but including also a space-time equaliser and a turbo decoder, the computational complexity of the detector grows exponentially both with the number of transmit antennas and with the modulation scheme. The APP (a posteriori probability) detector essentially compares the despread and pre-whitened received signal vector with all possible candidates (all possible symbol combinations from all transmitter antennas). Then the APP detector calculates soft outputs for the most likely transmitted symbol vector in the form of log-likelihood ratios (LLRs). With increasing numbers of transmitter antennas and modulation orders the number of possible candidates for the transmitted symbol vector, and hence the computational complexity, grows exponentially (2^{N_T*M} states with N_T transmitter antennas and M bits per symbol). This exponential growth in complexity makes implementations for MIMO with high-order modulations impractical (such as the case of four transmit and four receive antennas (4x4 antennas) using a 16-QAM or 64-QAM or higher-order Quadrature Amplitude Modulation (QAM) schemes).

Furthermore, the computational complexity of a MIMO detector has a significant effect on both the area (and therefore price) of the integrated circuit that would include the MIMO detector, and also its power consumption (which relates to battery lifetime). These characteristics are important, especially for high speed transmissions to the user equipment in MIMO HSDPA (Multiple-Input Multiple Output - High Speed Downlink Packet Access mode) for UMTS.

Summary of the Invention

An example of the present invention is a method of communication of data in a mobile telecommunications network involving at a transmitter first grouping data into a first sequence of bits and a second sequence of bits. There is then a step of modulating a signal with the bits of the first sequence so that the bits of the first sequence have a first level of communication error protection provided by the modulation and with the bits of the second sequence so that the bits of the second sequence have a second level of communication error protection provided by the modulation less than the first level of communication error protection. The signal is then transmitted. At a receiver, estimates of the bits of the first sequence from the signal are detected and contributions to the signal corresponding to the estimates are determined and cancelled from the signal so as to produce a modified signal. Estimates of the bits of the second sequence are then detected from the modified signal.

In some embodiments, at the transmitter, to handle higher order modulations, bit groups are encoded dependent on the level of protection provided by the modulation scheme. Bits which are to be given equivalent protection by the modulation scheme are encoded together in one block. In this way, in the receiver, the well-protected bits can be detected and their interference cancelled independently of the less-protected bits. Each data stream is detected (including being decoded) separately as 4-QAM symbols, and therefore with low computational complexity, even when the transmitted modulation scheme is 16-QAM, 64-QAM, 256-QAM or higher. This is achievable without loss of performance, in terms of bit error rate (BER) and frame error rate (FER).

In MIMO systems, this approach avoids the problem of known approaches of exponential growth in detector complexity with higher order modulation schemes such as 16-QAM and 64-QAM.

Brief Description of the Drawings

An exampleembodiment of the present invention will now be described with reference to the drawings, in which:

Figure 1 is a diagram illustrating cancellation of single bits from a 16QAM constellation,

Figure 2 is a diagram illustrating receiving circuitry to receive signals subject to Layered Encoding,

Figure 3 is diagram illustrating the receiver of the receiving circuitry shown in Figure 2,

Figure 4 is a diagram illustrating a MS-PPIC detector,

Figure 5 is a diagram illustrating 16-QAM modulation as an aggregate of 2 interdependent 4-QAM modulations,

Figure 6 is a diagram illustrating a transversal filter which is part of an equaliser, and

Figure 7 is a diagram illustrating selection of coefficients for the equaliser.

Detailed Description

In a 4 Quadrature Amplitude Modulation or 4 Quadrature Phase Shift Keying modulation scheme, bits corresponding to each symbol are allocated the same amount of energy and are therefore given the same amount of protection by the modulation scheme. In higher order modulation schemes such as 16-QAM, 64-QAM or 256-QAM, the modulated bits are not equally protected. The inventors realised that this fact can be made use of to introduce a layered encoding scheme, whereby bits which are given equivalent protection by the modulation scheme are encoded together in one block.

This allows us to first detect and decode the bit blocks which are well-protected by the modulation scheme, and subsequently subtract their contribution from the received signal in order to reduce the interference for the remaining less-protected bit blocks.

In this way, the received 16/64/256-QAM modulated signal can be treated as the sum of separately encoded 4-QAM data-streams which can be detected sequentially with any 4-QAM detection algorithm. Therefore even very high-order modulations like 256-QAM become feasible, since the computational complexity per information bit stays constant and does not grow exponentially as in the known receivers.

Figure 1 illustrates the process of bit-cancellation from a 16-QAM modulated symbol (which of course has four bits $b_{k,0}^{(n)}$, $b_{k,1}^{(n)}$, $b_{k,2}^{(n)}$, $b_{k,3}^{(n)}$). In this case bits $b_{k,0}^{(n)}$ & $b_{k,1}^{(n)}$, of each symbol are the most reliable bits and would be encoded as one block, i.e. bit stream. The remaining bits $b_{k,2}^{(n)}$ & $b_{k,3}^{(n)}$ of each symbol would be encoded as a separate lower reliability bit stream.

The basic detection process for 16-QAM would work as follows:

1..Detect high reliability bit stream (bits b1 & b2 of 16-QAM)

2. Calculate & cancel interference of high reliability bit stream $\Rightarrow$ reduce 16-QAM to 4-QAM

3. Detect low reliability bit stream (bits b3 & b4 of 16-QAM)

MIMO transmission

Figure 2 illustrates the system overview for the multiple-input multiple-output (MIMO) link with 16-QAM modulation, including transmitter and receiving circuitry.

At the transmitter 2, user data is encoded in encoders 4,6 using layered encoding scheme as described below, and then interleaved by interleavers 8,10 . The coded data stream is de-multiplexed into N_T sub-streams, corresponding to the N_T transmit antennas. Each sub-stream is then modulated by a 16QAM modulator 12 on to NK 16-QAM symbols and subsequently spread by spreading stage 14 by a factor Q via a set of K orthogonal spreading codes prior to transmission by transmit antennas 16. Each transmitted spread stream then occupies N symbol intervals. Also note that the same set of K codes are re-used across all transmit antennas. Therefore, the MIMO propagation environment, which is assumed to exhibit significant multipath, plays a major role in achieving signal separation by receiving circuitry 18.

Layered Encoding at the Transmitter

For a so-called Gray-mapped 16-QAM constellation, each symbol $x_k^{(n)}(t)$ is given by

$$x_k^{(n)}(t) = 2\{-b_{k,0}^{(n)}(t) - jb_{k,1}^{(n)}(t)\} + \{-b_{k,0}^{(n)}(t)b_{k,2}^{(n)}(t) - jb_{k,1}^{(n)}(t)b_{k,3}^{(n)}(t)\} \quad (1)$$

as a function of encoded bits $b_{k,0}^{(n)}$, $b_{k,1}^{(n)}$, $b_{k,2}^{(n)}$, $b_{k,3}^{(n)} \in \{-1, +1\}$. The corresponding constellation is illustrated in Figure 5(a). As can be seen, for such high-order constellations, the Euclidean distance is not the same for all modulated bits. This implies that the modulation scheme affords different levels of protection to different bits. For the Gray mapped 16-QAM constellation of Figure 5, it is clear that $b_{k,0}^{(n)}$ and $b_{k,1}^{(n)}$ are equally better protected than $b_{k,2}^{(n)}$ and $b_{k,3}^{(n)}$.

The feature of layered encoding is exploited by the receiving circuitry 18, whereby the well-protected bits $b_{k,0}^{(n)}(t)$ and $b_{k,1}^{(n)}(t)$ are detected and decoded first. Due to the greater Euclidean distance associated with these bits, they can be estimated reliably using a 4-QAM detector which is part of a 4-QAM receiver 20, treating the signal contributions from the remaining bits as interference. The contribution of the estimated bits is subsequently cancelled from the received signal. This significantly reduces the interference for the remaining less-protected bits $b_{k,2}^{(n)}(t)$ and $b_{k,3}^{(n)}(t)$, which are only then detected and decoded.

In order for the well-protected and less-protected bits to be detected and decoded separately, it is required that they are also encoded separately at the transmitter 2. This is indicated in Figure 2, where the user data is split into two classes and encoded/interleaved independently. The encoded bits of class-1 correspond to $b_{k,0}^{(n)}(t)$ and $b_{k,1}^{(n)}(t)$, while the encoded bits of class-2 correspond to $b_{k,2}^{(n)}(t)$ and $b_{k,3}^{(n)}(t)$. The bits are then mapped on to 16-QAM symbols according to Equation (1). For 64-QAM, the procedure is essentially the same, except that three classes are considered, according to the three levels of protection provided by the modulation scheme; for 256-QAM four classes are considered, and so on.

In an alternative but otherwise similar embodiment (not shown) to the example embodiment, the performance of the layered encoding scheme is further improved by the encoding rate of each sequence being adapted to the method of detection and channel conditions, for example by puncturing or repetition of bits in the coded sequence. In this way, forward error correction coding is adjusted for each sequence, i.e. layer, so as effect a trade-off between protecting subsequent layers and minimising the error propagation from previous layers. By doing this the bit-error rate of the receiver can be improved without altering the average code rate for a transmitted data block.

We now return to describing the example embodiment.

MIMO Reception

The transmitted signals are received by N_R receive antennas 22 after propagation through dispersive radio channels 24 with impulse response lengths of W chips. The received signal vector observed over the t^{th} symbol interval may then be written as

$$\begin{bmatrix} {}^{(1)}\underline{r} \\ \vdots \\ {}^{(N_R)}\underline{r} \end{bmatrix} = \begin{bmatrix} {}^{(1)}\boldsymbol{H}^{(1)} & \cdots & {}^{(1)}\boldsymbol{H}^{(N_T)} \\ \vdots & \ddots & \vdots \\ {}^{(N_R)}\boldsymbol{H}^{(1)} & \cdots & {}^{(N_R)}\boldsymbol{H}^{(N_T)} \end{bmatrix} \sum_{k=1}^{K} \boldsymbol{C}_k \begin{bmatrix} \underline{x}_k^{(1)} \\ \vdots \\ \underline{x}_k^{(N_T)} \end{bmatrix} + \begin{bmatrix} {}^{(1)}\underline{n} \\ \vdots \\ {}^{(N_R)}\underline{n} \end{bmatrix} \quad (2)$$

or

$$\underline{r} = \boldsymbol{H} \sum_{k=1}^{K} \boldsymbol{C}_k \underline{x}_k + \underline{n} \quad (3)$$

where ${}^{(m)}\underline{r} \in \mathbb{C}^{(QN+W-1)\times 1}$ is the signal received at the m^{th} antenna, ${}^{(m)}\boldsymbol{H}^{(i)} \in \mathbb{C}^{(QN+W-1)\times QN}$ is the channel matrix from the i^{th} transmit antenna to the m^{th} receive antenna, $\underline{x}_k^{(n)} \in \mathbb{C}^{N\times 1}$ is the symbol sequence $[x_k^{(n)}(1) \ldots x_k^{(n)}(N)]^T$ transmitted from the n^{th} antenna via the k^{th} spreading code, $\underline{n} \in \mathbb{C}^{(QN+W-1)\times 1}$ is a vector of i.i.d. zero-mean complex Gaussian random variables (i.e. $\boldsymbol{R}_n = \mathrm{E}\{\underline{n}\underline{n}^H\} = N_0\boldsymbol{I}$) representing

noise and inter-cell interference, and finally C_k is the spreading matrix for k^{th} spreading code, $\underline{c}_k \in \mathbb{C}^{Q\times 1}$, such that

$$C_k = \underbrace{\begin{bmatrix} \underline{c}_k & \cdots & \underline{0} \\ \vdots & \ddots & \vdots \\ \underline{0} & \cdots & \underline{c}_k \end{bmatrix}}_{N_T N \text{ Times}} \in \mathbb{C}^{QN_T N \times N_T N} \qquad (4)$$

The signal vector $\underline{r}$ is first applied to a processing stage 26 including a channel equalizer, de-spreader, and pre-whitener, then passed to the receiver 20.

As shown in Figure 3, the soft outputs computed by a detector 28 in the receiver 20 are then deinterleaved by a deinterleaver 30 and applied to a turbo decoder 32 also in the receiver 20. The turbo decoder 32 generates reliable estimates of the information bits, which are provided to output 34, and estimates of all the transmitted bits, which are provided to signal reconstruction stage 36 .

Receiver Circuitry

Figure 2 discussed above shows the receiver circuitry 18 which exploits the layered encoding scheme for the case of 16-QAM. The layered encoding scheme in conjunction with the 16-QAM transmitter 2 described in the previous section allows the receiving circuitry 18 to treat the transmitted symbols as the aggregate of two inter-dependent 4-QAM constellations. Bits $b_{k,0}^{(n)}$ and $b_{k,1}^{(n)}$ contribute to the first 4-QAM constellation, while bits $b_{k,2}^{(n)}$ and $b_{k,3}^{(n)}$ contribute to the second constellation (with the latter mapping depending on the values of $\{b_{k,0}^{(n)}, b_{k,1}^{(n)}\}$ for an overall Gray mapping). As shown in Figure 5, the 4-QAM receiver 20 first derives estimates of $\{b_{k,0}^{(n)}, b_{k,1}^{(n)}\}$ via detection and decoding, cancels their contribution from the received signal, and then derives estimates of $\{b_{k,2}^{(n)}, b_{k,3}^{(n)}\}$. The contributions to the signal due to the first bits and so corresponding to the estimates of the first bits are derived by modulating the bits as was undertaken at the transmitter and including the effect of the channel in known fashion and described in Equation 3 above but

without the noise term. It is clearly seen that once the contributions of $b_{k,0}^{(n)}$ and $b_{k,1}^{(n)}$ are subtracted from the 16-QAM constellation, the modulation is reduced to 4-QAM. In the particular example shown in Figure 5, the first two bits are estimated as $-1,+1$ (of course, giving bit values of 0,1). The cancellation of the first bits moves the remaining constellation points from the second quadrant (denoted Q2 in Figure 5(a)) to the centre, as shown in Figure 5(b). The remaining two bits are then estimated, in this case as $-1, -1$ (of course, giving bit values of 0,0).

While the layered receiver process has been described for 16-QAM, it can be readily extended to 64-QAM or higher orders, whereby the receiver treats the transmitted symbols as the aggregate of three or more inter-dependent 4-QAM constellations corresponding to three classes or more of reliability.

The proposed scheme can be used to demodulate data sent using a layered encoded high-order modulation scheme such as 16- or 64-QAM, using any type of low complexity 4-QAM detector. The layered encoding scheme can be used with receiving circuitry including known non-iterative (standard) or known iterative 4-QAM receivers 20.

Space-Time Equalization

If optimum space-time detection were used, it would imply joint detection of KN_T transmitted symbols per symbol epoch. For 4-QAM modulation, and for dispersive channels with intersymbol interference (ISI) extending over L symbols, this would require a search over a trellis containing $2^{2(L+1)KN_T}$ states. The computational complexity would be prohibitive for typical parameter values.

Note that, in flat fading conditions (L=0) and for K orthogonal codes re-used over the transmit antennas, the number of trellis states reduces to a more realistic value of 2^{2N_T}. Accordingly, an efficient strategy for dealing with dispersive (i.e. non-flat) channels is used of performing detection after a process of space-time equalization which effectively eliminates dispersion.

The equalization process in the equalizer of processing stage 26 inevitably causes noise colouring, which needs to be accounted for in the detection process.

The received signal over N symbol epochs is given by

$$\underline{r} = \boldsymbol{H}\sum_{k=1}^{K} \boldsymbol{C}_k \underline{x}_k + \underline{n} = \boldsymbol{HC}\underline{x} + \underline{n} = \boldsymbol{H}\underline{s} + \underline{n} \qquad (5)$$

where $\underline{s}=\boldsymbol{C}\underline{x}$ is the vector of spread symbols. A minimum mean-square error (MMSE) equalizer represents a space-time matrix V which minimizes the term $\mathrm{E}\{\|\underline{s}-V\underline{r}\|^2\}$. It is known that the solution to this problem is given by

$$V = \boldsymbol{R}_s \boldsymbol{H}^{\mathrm{H}} (\boldsymbol{H}\boldsymbol{R}_s \boldsymbol{H}^{\mathrm{H}} + \boldsymbol{R}_\nu)^{-1} \qquad (6)$$

where $\boldsymbol{R}_S=\mathrm{E}\{\underline{s}\underline{s}^{\mathrm{H}}\}=2\boldsymbol{CC}^{\mathrm{H}}$ since $\mathrm{E}\{\underline{x}\underline{x}^{\mathrm{H}}\}=2\boldsymbol{I}$ for 4-QAM. The equalization process may then be described as

$$\underline{e} = \boldsymbol{V}\underline{r} = \boldsymbol{VH}\sum_{k=1}^{K} \boldsymbol{C}_k \underline{x}_k + \boldsymbol{V}\underline{n} \quad \in \mathbb{C}^{QN_{\mathrm{T}}N} \qquad (7)$$

and clearly results in coloured noise. To avoid excessive computational complexity, space-time equalization is usually performed over a block of $N_{\mathrm{E}}<N$ symbol epochs and repeated N/N_{E} times to cover the entire transmission period. However, this reduction in complexity comes at the expense of degraded performance due to inaccuracies at the edges of the block.

De-spreading and Pre-whitening

The space-time equaliser removes most of the influence of the channel matrix H. As a result, assuming orthogonal spreading codes, the contribution of symbols transmitted using the k^{th} spreading code can be retrieved at the output of the equalizer via the de-spreading operation of the despreader which is part of processing stage 26.

Even with complete access to channel state information, the space time equalisation can never fully eliminate the influence of the MIMO channel (the zero-forcing equalizer achieves this at the expense of noise enhancement). In other words, $\boldsymbol{VH}=\boldsymbol{D}\neq\boldsymbol{I}$, where $\boldsymbol{D}$ is a non-diagonal distortion matrix.

This has a number of implications with respect to the computation of pre-whitened sufficient statistics for input to the detector, as described next. The output of the equalizer may be written as

$$\underline{e} = \boldsymbol{V}\underline{r} = \boldsymbol{VH}\sum_{k=1}^{K}\boldsymbol{C}_k\underline{x}_k + \boldsymbol{V}\underline{n} = \boldsymbol{D}\sum_{k=1}^{K}\boldsymbol{C}_k\underline{x}_k + \boldsymbol{V}\underline{n} \qquad (8)$$

and so the de-spreading operation for the k^{th} spreading code may be interpreted as

$$\begin{aligned}\underline{z}_k &= \|\underline{c}_k\|^{-2}\boldsymbol{C}_k^{\mathrm{H}}\underline{e} \\ &= \|\underline{c}_k\|^{-2}\boldsymbol{C}_k^{\mathrm{H}}\boldsymbol{D}\boldsymbol{C}_k\underline{x}_k + \|\underline{c}_k\|^{-2}\boldsymbol{C}_k^{\mathrm{H}}\boldsymbol{D}\boldsymbol{C}_{\mathrm{I},k}\underline{x}_{\mathrm{I},k} + \|\underline{c}_k\|^{-2}\boldsymbol{C}_k^{\mathrm{H}}\boldsymbol{V}\underline{n} \\ &= \boldsymbol{G}_k\underline{x}_k + \boldsymbol{T}_{\mathrm{I},k}\underline{x}_{\mathrm{I},k} + \boldsymbol{T}_k\underline{n} \\ &= \boldsymbol{G}_k\underline{x}_k + \underline{v}_{\mathrm{I},k} + \underline{v}_k \quad \in \mathrm{C}^{N_{\mathrm{T}}N} \qquad (9)\end{aligned}$$

where $\boldsymbol{C}_{\mathrm{I},k} \in \mathrm{C}^{QN_{\mathrm{T}}N \times N_{\mathrm{T}}N(K-1)}$ and $\underline{x}_{\mathrm{I},k} \in \mathrm{C}^{N_{\mathrm{T}}N(K-1)}$ are simply equal to the spreading matrix $\boldsymbol{C}$ and symbol vector $\underline{x}$ respectively with the elements associated with the k^{th} spreading code removed. The subscript '$_{\mathrm{I}}$' represents interference. Vector $\underline{z}_k$ consists of the equalized and de-spread contributions of $N_{\mathrm{T}}N$ symbols transmitted via the k^{th} spreading code over a total of N symbol epochs.

Considering only the N_{T} rows of Eq. (8) corresponding to the t^{th} symbol epoch, we have for $t = 1 \ldots N$

$$\begin{aligned}\underline{z}_k(t) &= \boldsymbol{G}_k(t)\underline{x}_k + \boldsymbol{T}_{\mathrm{I},k}(t)\underline{x}_{\mathrm{I},k} + \boldsymbol{T}_k(t)\underline{n} \\ &= \boldsymbol{B}_k(t)\underline{x}_k(t) + \widetilde{\boldsymbol{B}}_k(t)\underline{\widetilde{x}}_k(t) \\ &\quad + \boldsymbol{T}_{\mathrm{I},k}(t)\underline{x}_{\mathrm{I},k} + \boldsymbol{T}_k(t)\underline{n} \\ &= \boldsymbol{B}_k(t)\underline{x}_k(t) + \underline{s}_{\mathrm{I},k}(t) + \underline{v}_{\mathrm{I},k}(t) + \underline{v}_k(t) \\ &= \boldsymbol{B}_k(t)\underline{x}_k(t) + \underline{u}_k(t) \quad \in \mathrm{C}^{N_{\mathrm{T}}} \qquad (10)\end{aligned}$$

where $\underline{x}_k(t) \in C^{N_{\mathrm{T}}}$ is the vector of symbols transmitted during the t^{th} epoch while $\underline{\widetilde{x}}_k(t) \in C^{N_{\mathrm{T}}(N-1)}$ is the vector of symbols <u>not</u> transmitted during the t^{th} epoch via the

k^{th} spreading code. Note that while $\boldsymbol{B}_k(t)$ represents (spatial) self-interference, $\underline{s}_{\mathrm{I},k}(t)$ identifies space-time interference at the de-spreader output due to symbols transmitted via the k^{th} spreading code but at other symbol epochs. The imperfect operation of the space-time equalizer also implies that in addition to coloured noise, $\underline{v}_k$, a certain amount of coloured interference, $\underline{v}_{\mathrm{I},k}$, (originating from other spreading codes) also "leaks" through to the de-spreader output. Assuming that noise and interference are independent, one may write

$$\boldsymbol{R}_{\underline{u}_k(t)} = \mathrm{E}\{\underline{u}_k(t)\underline{u}_k^{\mathrm{H}}(t)\} \qquad (10)$$

$$= 2\{\widetilde{\boldsymbol{B}}_k(t)\widetilde{\boldsymbol{B}}_k^{\mathrm{H}}(t) + \boldsymbol{T}_{\mathrm{I},k}(t)\boldsymbol{T}_{\mathrm{I},k}^{\mathrm{H}}(t)\} + N_o\boldsymbol{T}_k(t)\boldsymbol{T}_k^{\mathrm{H}}(t)$$

since $\mathrm{E}\{\underline{x}_{\mathrm{I},k}\underline{x}_{\mathrm{I},k}^{\mathrm{H}}\} = 2\boldsymbol{I}_{N_\mathrm{T}N(K-1)}$ and $\mathrm{E}\{\underline{\widetilde{x}}_k(t)\underline{\widetilde{x}}_k^{\mathrm{H}}(t)\} = 2\boldsymbol{I}_{N_\mathrm{T}(N-1)}$.

Accordingly, the pre-whitening with respect to interference and noise is

$$\underline{z}_{\mathrm{w},k}(t) = \boldsymbol{R}_{\underline{u}_k(t)}^{-\frac{1}{2}}\underline{z}_k(t) = \boldsymbol{R}_{\underline{u}_k(t)}^{-\frac{1}{2}}\boldsymbol{B}_k(t)\underline{x}_k(t) + \boldsymbol{R}_{\underline{u}_k(t)}^{-\frac{1}{2}}\underline{u}_k(t)$$

$$= \boldsymbol{R}_{\underline{u}_k(t)}^{-\frac{1}{2}}\boldsymbol{B}_k(t)\underline{x}_k(t) + \underline{\varepsilon}_k(t) \qquad (12)$$

where $\mathrm{E}\{\underline{\varepsilon}_k(t)\underline{\varepsilon}_k^{\mathrm{H}}(t)\} = \boldsymbol{I}_{N_\mathrm{T}}$.

This pre-whitening function is performed by the pre-whitener which is part of processing stage 26.

Transversal Filter for Equalization

To avoid inaccuracies at block edges the matrix equaliser described in above Equation (7) is implemented as a transversal filter.

The channel matrix H consists of $N_R x N_T$ sub-matrices, each of the form of a convolution matrix with the coefficients of the corresponding channel from transmitter antenna n_T to receiver antenna n_R. The property that the minimum mean square error (MMSE) equalizer matrix V also consists of convolution matrix type sub-matrices, which perform a filter operation in order to equalize each of the channels, is exploited to implement the equalizer using known transversal filters in which the

weight coefficients $\underline{w}$ for each of the channels are derived from the block equalizer sub-matrices ${}^{(m)}V^{(n)}$.

As shown in Figure 7, for a 16-tap equalizer, the coefficients ${}^{(1)}\underline{w}^{(1)}$ are obtained by selecting the (Q+W-1)/2th column of the equalizer sub-matrix${}^{(1)}V^{(1)}$ of equalizer matrix of size N_E =1 symbol, where Q denotes the spreading factor and W the channel length. The example of ${}^{(1)}V^{(1)}$ in Figure 7 shows that the strongest elements of ${}^{(1)}\underline{w}^{(1)}$ are located in the middle. With increasing distance from the diagonal of the sub-matrix, the coefficients of ${}^{(m)}\underline{w}^{(n)}$ the become smaller, and approach zero for a sufficient number of equalizer taps.

Using this method, the maximum number of tap coefficients obtainable is N_EQ. However, since the calculation of V includes a matrix inversion, increasing N_E is undesirable due to the high increase in computational complexity.

For the transversal equalizer, the equalized signal for each receiver antenna can be written as

$$^{n_r}\underline{e} = \sum_{n=1}^{N_R} \mathrm{conv}\left\{ {}^{(n)}\underline{r}, {}^{(n)}\underline{w}^{(n_r)} \right\} \qquad (13)$$

This operation is equivalent to the block equalization in Equation (7) for a block size over all N symbol epochs, assuming the number of taps of the filter are sufficient large, that the coefficient in upper right and lower left triangle of the matrix ${}^{(m)}V^{(n)}$ which are not covered by the transversal equalizer approach zero. This operation is also equivalent to that shown schematically in Figure 6.

For the calculation of the pre-whitening matrix, the matrix equalizer matrix V is modified to match exactly the transversal filter operation. Then, the de-spreading and pre-whitening operation are performed as for the block- based equalization in Equations (8) – (12).

Approximate modelling of the equalizer output

Since the equalizer effectively eliminates the channel dispersion, the remaining intersymbol interference (ISI), which leaks from each symbol in the next, is relatively small in comparison to the distortion from the remaining. Therefore, the contribution

from other symbols to the sufficient statistics for the transmitter input is neglected and the N_T rows of Eq. (8) corresponding to the t^{th} symbol epoch are written as

$$\begin{aligned} \underline{z}_k(t) &\approx \boldsymbol{B}_k(t)\underline{x}_k(t) + \boldsymbol{T}_{1,k}(t)\underline{x}_{1,k} + \boldsymbol{T}_k(t)\underline{n} \\ &= \boldsymbol{B}_k(t)\underline{x}_k(t) + \underline{v}_{1,k}(t) + \underline{v}_k(t) \end{aligned} \qquad (14)$$

where $\underline{v}_{kI,k}(t)$ is the remaining interference from the other spreading codes and $\underline{v}_k(t)$ is coloured noise. The resulting correlation of interference and noise is

$$\begin{aligned} \boldsymbol{R}_{\underline{v}_k}(t) &= \mathrm{E}\{\underline{x}_{1,k}(t)\underline{x}^{H}_{1,k}(t)\} + \mathrm{E}\{\underline{v}_k(t)\underline{v}^{H}_k(t)\} \\ &= 2\boldsymbol{T}_{1,k}(t)\boldsymbol{T}^{H}_{1,k}(t)\} + N_o\boldsymbol{T}_k(t)\boldsymbol{T}^{H}_k(t) \end{aligned} \qquad (15)$$

, and it is this which is used instead of Equation (10) to pre-whiten according to Equation (12).

The detector

One option as to the detector 28 to use in receiver 20 (see Figure 3) is to use a known APP detector. The APP detector is basically a maximum likelihood detector which generates soft outputs in form of LLRs (Log-Likelihood Ratios).

Another option is a low complexity detector, namely a MS-PPIC detector. This detector can offer similar performance as the APP detector, at only about 20% of the computational complexity. Despite its low complexity, a receiver including the MS-PPIC detector is able to outperform an APP based receiver in dispersive channels, and also in combination with the layered encoding scheme.

These two types of detectors are considered in turn below.

A Posteriori Probability (APP) Detector

Consider pre-whitened sufficient statistics of the form

$$\underline{z}_w = A\underline{x} + \underline{\varepsilon} \qquad (16)$$

where $\underline{x} \in \mathbb{C}^{N_T}$ is the vector of transmitted symbols and $A \in \mathbb{C}^{N_T \times N_T}$ is the transformation matrix. Under the assumption that the elements of the additive disturbance vector are independent identical distributed (i.i.d.) zero-mean complex

Gaussian random variables of unit variance (i.e. $\mathrm{E}\{\underline{\varepsilon}\underline{\varepsilon}^{H}\} = \boldsymbol{I}$), the likelihood function or conditional probability density of $\underline{z}_{\mathrm{w}}$ may be written as

$$\begin{aligned}
\mathrm{f}(\underline{z}_{\mathrm{w}} | \underline{x}) &= \prod_{i=1}^{N_{\mathrm{T}}} \mathrm{f}([\underline{z}_{\mathrm{w}}]_i | \underline{x}) \\
&= \prod_{i=1}^{N_{\mathrm{T}}} \frac{1}{\pi} \exp\left\{ -\left| [\underline{z}_{\mathrm{w}}]_i - [A\underline{x}]_i \right|^2 \right\} \\
&= \pi^{-N_{\mathrm{T}}} \exp\left\{ -\left\| \underline{z}_{\mathrm{w}} - A\underline{x} \right\|^2 \right\}
\end{aligned}$$

With the availability of sufficient statistics $\underline{z}_{\mathrm{w}}$, a detector is in a position to make a hypothesis $\underline{x}_O$ regarding the transmitted symbols. The probability that this hypothesis is correct is equal to the probability, $\mathrm{P}\{\underline{x}_O | \underline{z}_{\mathrm{w}}\}$, that $\underline{x}_O$ was indeed transmitted given $\underline{z}_{\mathrm{w}}$. The maximum *a posteriori* probability (MAP) detector is defined as that which minimizes the probability of an incorrect hypothesis:

$$\begin{aligned}
\hat{\underline{x}}_{\mathrm{MAP}} &= \underset{\underline{x}}{\operatorname{argmax}}\, \mathrm{P}\{\underline{x} | \underline{z}_{\mathrm{w}}\} = \underset{\underline{x}}{\operatorname{argmax}} \frac{\mathrm{P}\{\underline{x}, \underline{z}_{\mathrm{w}}\}}{\mathrm{f}(\underline{z}_{\mathrm{w}}) d\underline{z}_{\mathrm{w}}} \\
&= \underset{\underline{x}}{\operatorname{argmax}} \frac{\mathrm{f}(\underline{z}_{\mathrm{w}} | \underline{x}) d\underline{z}_{\mathrm{w}} \; \mathrm{P}\{\underline{x}\}}{\mathrm{f}(\underline{z}_{\mathrm{w}}) d\underline{z}_{\mathrm{w}}} \\
&= \underset{\underline{x}}{\operatorname{argmax}}\, \mathrm{f}(\underline{z}_{\mathrm{w}} | \underline{x}) \mathrm{P}\{\underline{x}\} \\
&= \underset{\underline{x}}{\operatorname{argmax}} \frac{\mathrm{P}\{\underline{x}\}}{\pi^{N_{\mathrm{T}}}} \exp\left\{ -\left\| \underline{z}_{\mathrm{w}} - A\underline{x} \right\|^2 \right\}
\end{aligned}$$

$$\hat{\underline{x}}_{\mathrm{MAP}} = \underset{\underline{x}}{\operatorname{argmin}} \left\{ \left\| \underline{z}_{\mathrm{w}} - A\underline{x} \right\|^2 - \ln(\mathrm{P}\{\underline{x}\}) \right\} \qquad (17)$$

where $\mathrm{P}\{\underline{x}\}$ is the *a priori* probability of $\underline{x}$.

$$\ln(\mathrm{P}\{\underline{x}\}) = -\frac{1}{2} \underline{b}^{\mathrm{T}} \Lambda_a(\underline{b}) \qquad (18)$$

In the absence of such *a priori* information, the MAP detector degenerates into the maximum likelihood (ML) detector.

Soft outputs for the i^{th} bit of the symbol vector $\underline{x}$ may be derived in the form of log-likelihood ratios (LLR) at the output of the MAP detector

$$\Lambda(b_i) = \ln\frac{P\{b_i = +1 \mid \underline{z}_w\}}{P\{b_i = -1 \mid \underline{z}_w\}} = \ln\frac{\sum_{\underline{x}\mid b_i=+1} P\{\underline{x} \mid \underline{z}_w\}}{\sum_{\underline{x}\mid b_i=-1} P\{\underline{x} \mid \underline{z}_w\}}$$

$$= \ln\frac{\sum_{\underline{x}\mid b_i=+1} f(\underline{z}_w \mid \underline{x}) P\{\underline{x}\}}{\sum_{\underline{x}\mid b_i=-1} f(\underline{z}_w \mid \underline{x}) P\{\underline{x}\}}$$

$$= \ln\frac{\sum_{\underline{x}\mid b_i=+1} \pi^{-N_T} \exp\left\{-\left\|\underline{z}_w - A\underline{x}\right\|^2\right\} P\{\underline{x}\}}{\sum_{\underline{x}\mid b_i=-1} \pi^{-N_T} \exp\left\{-\left\|\underline{z}_w - A\underline{x}\right\|^2\right\} P\{\underline{x}\}}$$

$$\Lambda(b_i) = \ln\frac{\sum_{\underline{x}\mid b_i=+1} \exp\left\{-\left\|\underline{z}_w - A\underline{x}\right\|^2 + \ln P\{\underline{x}\}\right\}}{\sum_{\underline{x}\mid b_i=-1} \exp\left\{-\left\|\underline{z}_w - A\underline{x}\right\|^2 + \ln P\{\underline{x}\}\right\}} \qquad (19)$$

Equation (19) represents what is commonly known as the *a posteriori* probability (APP) detector. Comparison of Eqs. (18) and (19) indicate that the signs of the above LLR values are equivalent to minimum probability of error (MAP) bit estimates.

As can be seen, the expression for the LLR is not computationally friendly and involves divisions, logarithms and exponentials. The computation of the LLR can be simplified by exploiting the max-log approximation which states that $\ln(e^{\delta_1} + e^{\delta_2} + \Lambda + e^{\delta_n}) \approx \max(\delta_1, \delta_2, \Lambda\ \delta_n)$. Then the max-log-APP detector may be written as:

$$\begin{aligned}\Lambda(b_i) &\approx \max_{\underline{x}\mid b_i=+1}\left\{-\left\|\underline{z}_w - A\underline{x}\right\|^2 + \ln P\{x\}\right\} \\ &\qquad - \max_{\underline{x}\mid b_i=-1}\left\{-\left\|\underline{z}_w - A\underline{x}\right\|^2 + \ln P\{x\}\right\} \\ &\approx \min_{\underline{x}\mid b_i=-1}\left\{\left\|\underline{z}_w - A\underline{x}\right\|^2 - \ln P\{x\}\right\} \\ &\qquad - \min_{\underline{x}\mid b_i=+1}\left\{\left\|\underline{z}_w - A\underline{x}\right\|^2 - \ln P\{x\}\right\}\end{aligned} \qquad (20)$$

Multi-Stage Parallel Interference Canceller

The multi-stage partial parallel interference canceller (MS-PPIC) detector is considered here as an alternative to APP-type detection in the context of MIMO downlink. The MS-PPIC detector is shown in Figure 4. It operates in an iterative manner, initialised by matched filter outputs (with or without channel equalizer) and generates high quality soft outputs based on the nonlinear cancellation behaviour.

Having computed the set of pre-whitened sufficient statistics $\underline{z}_{w,k}(t)$ for $k = 1 \ldots K$ and $t = 1 \ldots N$, these vectors can be individually applied to the detector. Consider

$$\underline{z}_{\mathrm{w}} = A\underline{x} + \underline{\varepsilon} \qquad (21)$$

where $\underline{x} \in \mathrm{C}^{N_\mathrm{T}}$ is the vector of transmitted symbols, $\boldsymbol{A} \in \mathrm{C}^{N_\mathrm{T} \times N_\mathrm{T}}$ is the transformation matrix and $\mathrm{E}\{\underline{\varepsilon\varepsilon}^{\mathrm{H}}\} = \boldsymbol{I}$. Performing matched filtering and normalizing we have

$$\begin{aligned} \underline{y} &= \Delta^{-1} A^{\mathrm{H}} \underline{z}_{\mathrm{w}} \\ &= \Delta^{-1} A^{\mathrm{H}} A \underline{x} + \Delta^{-1} A^{\mathrm{H}} \underline{\varepsilon} \qquad (22) \\ &= \Delta^{-1} \boldsymbol{R} \underline{x} + \underline{\eta} \end{aligned}$$

where $\boldsymbol{R}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$, $\Delta=\mathrm{diag}\{\boldsymbol{R}\}$ and $E\{\underline{\eta\eta}^{\mathrm{H}}\} = \Delta^{-1} \boldsymbol{R} \Delta^{-\mathrm{H}}$.

The matched filter output may then be written in the form

$$\begin{aligned} \underline{y} &= \underline{x} + \Delta^{-1}(\boldsymbol{R} - \Delta)\underline{x} + \underline{\eta} \\ &= \underline{x} + \Delta^{-1}\boldsymbol{R}'\underline{x} + \underline{\eta} \qquad (23) \\ &= \underline{x} + \boldsymbol{S}\underline{x} + \underline{\eta} \end{aligned}$$

where, given that $\boldsymbol{R}'$ and $\boldsymbol{S}$ both have zero diagonals, it is clear that the term $\boldsymbol{S}\underline{x}$ represents the interference contributions which need to be cancelled. The sufficient

statistics of Eq. (5.3) are input to the MS-PPIC and may be viewed as the 0th stage output of the detector. Denoting the n^{th} element of $\underline{y}$ as $y^{(n)}$ and the n^{th} row of $\boldsymbol{S}$ as $\underline{s}^{(n)\mathrm{H}}$, we then have

$$\begin{aligned} y^{(n)}[0] &= y^{(n)} \\ &= x^{(n)} + \underline{s}^{(n)\mathrm{H}}\,\underline{x} + \eta^{(n)} \qquad (24) \\ &= x^{(n)} + v^{(n)}[0] \end{aligned}$$

and it immediately follows that cancellation at the m^{th} stage of the detector should be of the form

$$\begin{aligned} y^{(n)}[m] &= y^{(n)}[0] - \underline{s}^{(n)\mathrm{H}} \mathrm{f}\{\underline{\hat{x}}[m-1]\} \\ &= x^{(n)} + \underline{s}^{(n)\mathrm{H}}(\underline{x} - \mathrm{f}\{\underline{\hat{x}}[m-1]\}) + \eta^{(n)} \qquad (25) \\ &= x^{(n)} + v^{(n)}[m] \end{aligned}$$

where $\mathrm{f}\{\underline{\hat{x}}[m-1]\}$ is in general a non-linear function of tentative estimates, $\underline{\hat{x}}[m-1]$, derived in the previous stage. This is illustrated schematically in Figure 4.

One could ignore the non-linearity and simply use the tentative estimates $\underline{\hat{x}}[m-1]$ directly in a linear cancellation process. It has been shown that (under certain constraints on the eigenvalues of $\boldsymbol{S}$) the resulting linear MS-PIC converges to the MMSE joint-detector as the number of stages approaches infinity []. At the other extreme, one could choose the function $\mathrm{f}\{\bullet\}$ to be a mapping to the 4-QAM alphabet (i.e. a threshold operation). Such hard cancellation would perform well if and only if there was a high level of confidence regarding the reliability of tentative estimates $\underline{\hat{x}}[m-1]$.

In order to deal with cases where the tentative estimates are unreliable, one may instead use the expected value of the tentative estimates $\underline{\hat{x}}[m-1]$ in the cancellation process.

Since $y^{(n)}[m-1] = x^{(n)} + v^{(n)}[m-1]$

then $\hat{x}^{(n)}[m-1] = y^{(n)}[m-1]$

and assuming that the noise+interference term $\nu^{(n)}$ is Gaussian distributed, it can readily be shown that

$$
\begin{aligned}
f\left\{\underline{\hat{x}}^{(n)}[m-1]\right\} &\equiv E\left\{\mathrm{Re}(\underline{\hat{x}}^{(n)}[m-1])\right\} + jE\left\{\mathrm{Im}(\underline{\hat{x}}^{(n)}[m-1])\right\} \\
&\equiv E\left\{\underline{\hat{b}}_0^{(n)}[m-1]\right\} + jE\left\{\underline{\hat{b}}_1^{(n)}[m-1]\right\} \\
&\equiv \tanh\left\{\tfrac{1}{2}\Lambda(\underline{\hat{b}}_0^{(n)}[m-1])\right\} + j\tanh\left\{\tfrac{1}{2}\Lambda(\underline{\hat{b}}_1^{(n)}[m-1])\right\}
\end{aligned}
$$

$$
\begin{aligned}
&\equiv \tanh\left\{\alpha^{(n)}[m]\ \mathrm{Re}(\underline{y}^{(n)}[m-1])\right\} \\
&\quad + j\tanh\left\{\alpha^{(n)}[m]\ \mathrm{Im}(\underline{y}^{(n)}[m-1])\right\}
\end{aligned}
\qquad (26)
$$

where $\Lambda(\bullet)$ is the log-likelihood ratio and

$$\alpha^{(n)}[m] = \frac{2}{\sigma^2_{\nu^{(n)}[m-1]}} \qquad (27)$$

can be viewed as an antenna-dependent "softness" factor for the m^{th} stage. As can be seen from (5.4), $\alpha^{(n)}[m]$ can be readily computed for the first stage:

$$\alpha^{(n)}[1] = \frac{2}{E\left\{\left|\underline{s}^{(n)H}\underline{x} + \eta^{(n)}\right|^2\right\}} = \frac{2}{2\underline{s}^{(n)H}\underline{s}^{(n)} + R^{-1}_{n,n}} \qquad (28)$$

with $R_{n,n}$ the n^{th} diagonal element of $\boldsymbol{R}$. The computation of $\alpha^{(n)}[m]$ is more involved for subsequent stages. Consequently, $\alpha^{(n)}[1]$ may be used for all stages m=1…M. Though sub-optimal, this strategy should not significantly degrade performance in the SNR range of interest.

Finally, the M stages of parallel cancellation may be described as

$$
\begin{array}{l}
\text{for } m = 1\Lambda\ M \ \text{(stages)} \\
\quad \underline{\xi} = \underline{y}[m-1] \in \mathbb{C}^{N_T} \\
\quad \text{for } n = 1\Lambda\ N_T \ \text{(antennas)} \\
\qquad y^{(n)}[m] = y^{(n)}[0] \\
\qquad\qquad - s^{(n)H}\left\{\tanh\{\Gamma\,\mathrm{Re}(\underline{\xi})\} + j\tanh\{\Gamma\,\mathrm{Im}(\underline{\xi})\}\right\} \\
\qquad \xi^{(n)} = y^{(n)}[m] \\
\quad \text{end} \\
\text{end}
\end{array}
$$

where $\Gamma = 2\left[diag\{2SS^{H}\}+\Delta^{-1}\right]^{-1}$ (29)

is a diagonal matrix of the "softness" factors. Essentially, at each stage the contributions due to other antennas are removed from the elements of $\underline{y}[0]$. The contributions at the m^{th} stage are constructed from "soft symbols" derived in the previous $(m-1)^{th}$ stage as well as those derived most recently in the current stage. Log-likelihood ratios may be computed after the last stage, where as a result of multiple stages of cancellation $y^{(n)}[M] \sim x^{(n)} + \eta^{(n)}$ and so

$$\Lambda(b_0^{(n)}) = \frac{4\,\mathrm{Re}(y^{(n)}[M])}{R^{-1}_{n,n}} \qquad \Lambda(b_1^{(n)}) = \frac{4\,\mathrm{Im}(y^{(n)}[M])}{R^{-1}_{n,n}} \qquad (30)$$

Complexity comparison

Modulation Scheme	Original APP	APP + Layered Encoding	MS-PPIC + Layered Encoding
4-QAM	2048	2048	544
16-QAM	$524*10^3$	2048	544
64-QAM	$134*10^6$	2048	544
256-QAM	$34*10^9$	2048	544

TABLE 1: COMPLEXITY COMPARISON (Scenario: 4x4 Antennas, 1/3 rate coding)

Table 1 shows a complexity comparison in multiplications per symbol period between comparative examples of a known receiver including an APP detector and the two proposed schemes based on reception of layered encoding (involving an APP detector and an MS-PPIC detector respectively). Each is considered in a scenario where there are 4 transmit antennas, 4 receive antennas and 1 bit of data becomes 3 encoded bits including error check data (denoted 1/3 rate coding). The computational complexity in the case of the known receiver including an APP detector (denoted "original APP" in the Table) grows exponentially. Therefore, when high-order modulations are used, the complexity becomes clearly prohibitive. On the other hand, it will be seen that with the proposed reception of layered encoding, the complexity

per information bit stays constant for all modulations schemes. Additionally, the proposed scheme involving the MS-PPIC based detector reduces the complexity by a further 75% and allows high-speed MIMO receivers, capable of dealing with even 256-QAM modulation at very low computational complexity.

It is seen from the table that the proposed reception of layered encoding can have particular advantages in avoiding the exponential growth in complexity that occurs in known APP based receivers using higher order modulation. The receiver based on the APP detector and reception of layered encoding has an advantage that existing MIMO chips, can be reused to provide extremely high modulation schemes for MIMO HSDPA.

The receiver based on a MS-PPIC detector and reception of layered encoding has an advantage that computational complexity of the MS-PPIC detector is only 20% of the known APP-based receiver, and can achieve even better performance.

The reception of layered encoding scheme is not restricted to these two types of detectors, but can be used in conjunction with any 4-QAM capable detector.

Exploiting the layered encoding scheme in the proposed receivers (as described above) allows the use of higher order modulations (16-, 64-, 256-QAM) without exponential increase in computational complexity whilst maintaining good bit error rate/frame error rate (BER/FER) performance.